Achievable Rates for the Broadcast Channel with Feedback

Ramji Venkataramanan
Dept. of Electrical Engineering
Stanford University, USA
Email: vramji@stanford.edu

S. Sandeep Pradhan
Dept. of Electrical Engineering & Computer Science
University of Michigan, Ann Arbor, USA
Email: pradhanv@eecs.umich.edu

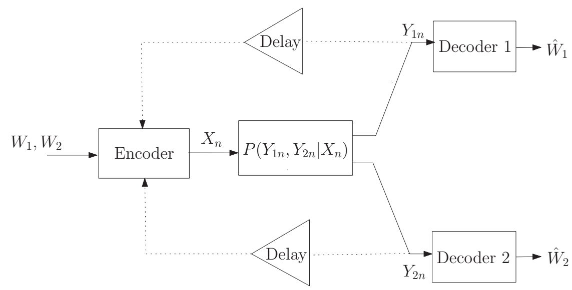

Fig. 1. The discrete memoryless broadcast channel with feedback.

***Abstract*—A single-letter achievable rate region is proposed for the two-receiver discrete memoryless broadcast channel with feedback. It is shown through an example that the rate-region can be strictly larger than the no-feedback capacity region. The coding strategy involves block-Markov superposition coding using Marton's scheme as the starting point. If the message rates in the Marton scheme are too high to be decoded at the end of a block, each receiver is left with a list of messages compatible with its output. In the next block, we send resolution information for each receiver to resolve its list. The key observation is that the resolution information of the first receiver is correlated with that of the second. We transmit this correlated information efficiently in the following block using ideas from the Han-Costa coding scheme.**

I. Introduction

The two-receiver discrete memoryless broadcast channel is shown in Figure 1. The channel has input X, outputs Y_1, Y_2, and is characterized by the conditional law $P_{Y_1Y_2|X}$. The largest known set of achievable rates for this channel without feedback is due to Marton [1]. Marton's rate region is equal the capacity region in all cases where it is known. (See [2], for example, for a list of such channels.)

El Gamal showed in [3] that feedback does not enlarge the capacity region of a physically degraded broadcast channel. Later, through a simple example, Dueck [4] demonstrated that feedback can strictly improve the capacity region of a general broadcast channel. For the degraded AWGN broadcast channel with feedback, achievable rates larger than the no-feedback capacity were established in [5], and more recently, in [6]. In this paper, we establish a single-letter achievable rate region for the discrete memoryless broadcast channel with feedback.

This work was supported in part by NSF grants CCF-0448115 (CAREER), CCF-0915619.

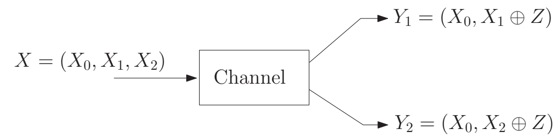

Fig. 2. The channel input is a binary triple (X_0, X_1, X_2). $Z \sim$ Bernoulli$(\frac{1}{2})$ is an independent noise variable.

Before describing our coding strategy, let us revisit the example from [4]. Consider the broadcast channel in Figure 2. The channel input is a binary triple (X_0, X_1, X_2). X_0 is transmitted cleanly to both receivers. In addition, receiver 1 receives $X_1 \oplus Z$ and receiver 2 receives $X_2 \oplus Z$, where Z is an independent binary Bernoulli$(\frac{1}{2})$ noise variable. In the above, the operation $\oplus$ denotes the modulo-two sum. Without feedback, the maximum sum rate for this channel is 1 bit/channel use, achieved by using the clean input X_0 alone. In other words, no information can be reliably transmitted through inputs X_1 and X_2.

Dueck described a simple scheme to achieve a greater sum rate using feedback. In the first channel use, transmit one bit to each receiver i through X_i, $i = 1, 2$. Receiver i then receives $Y_i = X_i \oplus Z$, and cannot recover X_i. The transmitter learns Y_1, Y_2 through feedback and can compute $Z = Y_1 \oplus X_1 = Y_2 \oplus X_2$. For the next channel use, the transmitter sets $X_0 = Z$. Since X_0 is received noiselessly by both receivers, receiver i can now recover X_i as $Y_i \oplus Z$. We can repeat this idea over several transmissions: in each channel use, transmit a fresh pair of bits (through X_1, X_2) *and* the noise realization of the previous channel use (through X_0). This yields a sum rate of 2 bits/channel use. This is, in fact, the sum-capacity of the channel since it equals the cut-set bound $\max_{P_X} I(X; Y_1Y_2)$.

The example suggests a natural way to exploit feedback in a broadcast channel. If we transmit a block of information at rates outside the no-feedback capacity region, the receivers cannot uniquely decode their messages at the end of the block. Each receiver now has a list of its codewords that are jointly typical with its channel output. In the next block, we attempt to resolve these lists at the two receivers. The key observation is that the resolution information needed by receiver 1 is in general *correlated* with the resolution information needed by receiver 2. The above example is an extreme case of this: the resolution information needed by the two receivers is identical, i.e., the correlation is perfect!

It is known that correlated information can be transmitted over the broadcast channel at higher rates than independent information [7]–[11]. At the heart of the proposed coding scheme is a way to represent the resolution information of the two receivers as a pair of correlated sources, which is then transmitted efficiently in the next block using the techiques of [7]. We repeat this idea over several blocks of transmission, with each block containing independent fresh information superimposed over correlated resolution information for the previous block.

II. PREMILINARIES

We use uppercase letters to denote random variables, lowercase for their realizations and calligraphic notation for their alphabets. Bold-face notation is used for random vectors. Unless otherwise stated, all vectors have length n. Thus $\mathbf{A} \triangleq A^n \triangleq (A_1, \dots, A_n)$. The typical set of a random variable with pmf P is denoted $T(P)$.

Consider a two-user discrete memoryless broadcast channel with input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$, defined by the conditional probability mass function $P_{Y_1 Y_2|X}(.|x)$ for all $x \in \mathcal{X}$. The channel is memoryless, i.e., it satisfies

$$P(Y_{1i}, Y_{2i}|X^i, Y_1^{i-1}, Y_2^{i-1}) = P(Y_{1i}, Y_{2i}|X_i),\ i = 1, 2, \dots$$

There is feedback from both receivers to the transmitter.

Definition 1: An $(n, 2^{NR_1}, 2^{nR_2})$ code with block length n and rates (R_1, R_2) for a broadcast channel with feedback consists of

1) A sequence of encoder mappings:

$$e_i : \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\} \times \mathcal{Y}_1^{i-1} \times \mathcal{Y}_2^{i-1} \to \mathcal{X}, \quad i = 1, \dots, n$$

2) Decoder mappings given by

$$g_1 : \mathcal{Y}_1^n \to \{1, \dots, 2^{nR_1}\},$$
$$g_2 : \mathcal{Y}_2^n \to \{1, \dots, 2^{nR_2}\}.$$

Assuming the messages (W_1, W_2) are drawn uniformly from the set $\{1, \dots 2^{nR_1}\} \times \{1, \dots 2^{nR_2}\}$, the error probability is

$$P_{e,n} = \frac{\sum_{k,l} \Pr(g_1(\mathbf{Y}_1) \neq k \text{ or } g_2(\mathbf{Y}_2) \neq l | W_1, W_2 = k, l)}{2^{nR_1} \cdot 2^{nR_2}}$$

A rate pair (R_1, R_2) is achievable for the broadcast channel with feedback if there exists a sequence of $(n, 2^{nR_1}, 2^{nR_2})$ codes such that $P_{e,n} \to 0$ as $n \to \infty$. The closure of all achievable rate pairs is the capacity region with feedback.

III. MAIN RESULT

A. Overview of Coding Scheme

Consider B blocks of transmission, with a fresh pair of messages (m_{1b}, m_{2b}) in each block b $(1 \leq b \leq B)$. Let the message rate pair (R_1, R_2) lie outside Marton's region [1].

In each block, the message pair is encoded using Marton coding. Random variables U and V carry the messages for receivers 1 and 2, respectively.[1] $\mathbf{U}$ and $\mathbf{V}$ codebooks are chosen according to P_U and P_V, and are divided into 2^{nR_1} and 2^{nR_2} bins, respectively. To encode message pair (m_1, m_2), we choose one codeword from bin m_{1b} of the U-codebook and one codeword from bin m_{2b} of the V-codebook that are jointly typical according to P_{UV}. This pair of jointly typical codewords is set to be $\mathbf{U}_b, \mathbf{V}_b$. The encoding is successful if each bin has is large enough [12].

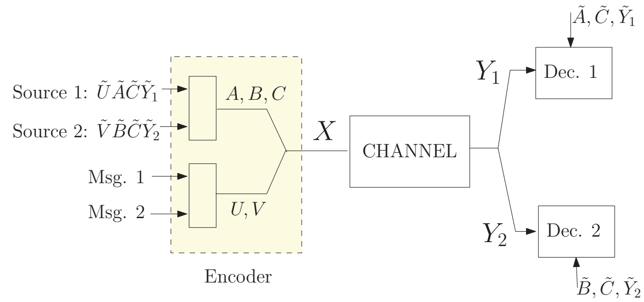

Fig. 3. Transmitting correlated sources with side-information at the receivers through (A, B, C), and fresh information through U, V.

However, the sizes of the U and V codebooks are too large for receivers 1 and 2 to uniquely decode $\mathbf{U}_b$ and $\mathbf{V}_b$, respectively. Hence receiver 1 is left with a list of U-codewords that are jointly typical with $\mathbf{Y}_{1b}$; receiver 2 has a similar list of V-codewords that are jointly typical with $\mathbf{Y}_{2b}$. The transmitter knows both these lists due to feedback, and resolves them in the next block block as follows. If $(\mathbf{U}_b, \mathbf{V}_b)$ was the actual codeword pair transmitted, then $(\mathbf{Y}_{1b}, \mathbf{U}_b)$ and $(\mathbf{Y}_{2b}, \mathbf{V}_b)$ may be considered realizations of a pair of correlated sources ($Y_1 U$ and $Y_2 V$), jointly distributed according to $P_{UVY_1Y_2}$. The goal in block $(b+1)$ is to transmit this correlated pair over the broadcast channel, with receiver 1 needing to decode $\mathbf{U}_b$, and receiver 2 decoding $\mathbf{V}_b$.

To transmit the pair of correlated sources, we denote them in block $(b+1)$ as

$$\tilde{\mathbf{U}}_{b+1} = \mathbf{U}_b,\ \tilde{\mathbf{Y}}_{1(b+1)} = \mathbf{Y}_{1b},\ \tilde{\mathbf{V}}_{b+1} = \mathbf{V}_b,\ \tilde{\mathbf{Y}}_{2(b+1)} = \mathbf{Y}_{2b}.$$

We then use the ideas of Han and Costa [7], [9] to transmit this pair of correlated sources $(\tilde{U}\tilde{Y}_1, \tilde{V}\tilde{Y}_2)$ over the broadcast channel. A correlated triple of random variables (A, B, C) is introduced to cover the sources. At the end of block $(b+1)$, receiver 1 determines $\tilde{U}$ by decoding (A, C). Since $\tilde{\mathbf{U}}_{b+1} = \mathbf{U}_b$, receiver 1 has thus decoded its message for block b with a *delay* of one block. Similarly, receiver 2 determines $\tilde{V}$ by decoding (B, C), thereby decoding its message for block b.

As described above, the encoder generates two sets of variables in each block: (A, B, C) to cover the correlated sources, and (U, V) to represent the fresh messages. It combines them together to generate the channel input X. To get a single-letter characterization of achievable rates, we need to ensure that the random variables in each block follow a stationary

[1] A common random variable W may also be used, but the essence of the coding strategy is the same. A longer version of this paper will present a coding scheme that incorporates a common random variable.

joint distribution. We now describe how to ensure that the sequences in each block are jointly distributed according to

$$P_{ABC} \cdot P_{UV} \cdot P_{X|ABCUV} \cdot P_{Y_1 Y_2|X} \tag{1}$$

for some chosen P_{ABC}, P_{UV} and $P_{X|ABCUV}$. At the end of each block , receiver 1 decodes A and C from Y_1, but cannot decode U. These sequences, represented using $\sim$ notation, become the source to be transmitted to receiver 1 in the next block. This source is the entire tuple $\tilde{A}\tilde{C}\tilde{U}\tilde{Y}_1$, which receiver 1 decodes using $\tilde{A}\tilde{C}\tilde{Y}_1$ as side-information. This is shown in Figure 3. Similarly, $\tilde{B}\tilde{C}\tilde{V}\tilde{Y}_2$ is the source for receiver 2, which decodes it using $\tilde{B}\tilde{C}\tilde{Y}_2$ as side-information.

Suppose that the sequences in a given block are jointly distributed according to (1). These sequences become the source pair $(\tilde{A}\tilde{C}\tilde{U}\tilde{Y}_1, \tilde{B}\tilde{C}\tilde{V}\tilde{Y}_2)$ in the next block. To cover the source pair with A, B, C, we pick a conditional distribution

$$Q_{ABC|\tilde{A}\tilde{B}\tilde{C}\tilde{U}\tilde{V}\tilde{Y}_1\tilde{Y}_2}$$

such that the covering sequences are distributed according to P_{ABC}. This holds when the condition given by (3) below is satisfied. We thereby ensure that the sequences in *each* block are jointly distributed according to (1).

Our technique of exploiting the correlation induced by feedback is similar in spirit to the coding scheme of Han for two-way channels [13]. We now state the theorem and give a sketch of the proof in Section IV.

Theorem 1: For the broadcast channel described by $P_{Y_1,Y_2|X}$, fix any joint distribution of the form

$$P_{ABC} \cdot P_{UV} \cdot P_{X|ABCUV} \cdot P_{Y_1 Y_2|X} \tag{2}$$

where A, B, C, U, V are random variables defined over discrete alphabets. Let $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y}_1, \tilde{Y}_2)$ denote the corresponding variables for the previous block. Fix a conditional distribution $Q_{ABC|\tilde{A},\tilde{B},\tilde{C},\tilde{U},\tilde{V},\tilde{Y}_1,\tilde{Y}_2}$ such that

$$P_{ABC}(.) = \sum_{\tilde{a},\tilde{b},\tilde{c},\tilde{u},\tilde{v},\tilde{y}_1,\tilde{y}_2} Q_{ABC|\tilde{A}\tilde{B}\tilde{C}\tilde{U}\tilde{V}\tilde{Y}_1\tilde{Y}_2}(.|\tilde{a},\tilde{b},\tilde{c},\tilde{u},\tilde{v},\tilde{y}_1,\tilde{y}_2) \cdot P_{ABCUVY_1Y_2}(\tilde{a},\tilde{b},\tilde{c},\tilde{u},\tilde{v},\tilde{y}_1,\tilde{y}_2). \tag{3}$$

The joint distribution for two successive blocks is then

$$P \triangleq P_{\tilde{A}\tilde{B}\tilde{C}\tilde{U}\tilde{V}\tilde{X}\tilde{Y}_1\tilde{Y}_2} \cdot Q_{ABC|\tilde{A},\tilde{B},\tilde{C},\tilde{U},\tilde{V},\tilde{Y}_1,\tilde{Y}_2} \cdot P_{ABCUVXY_1Y_2} \tag{4}$$

where $P_{\tilde{A}\tilde{B}\tilde{C}\tilde{U}\tilde{V}\tilde{X}\tilde{Y}_1\tilde{Y}_2}$ and $P_{ABCUVXY_1Y_2}$ are both given by (2). For any such joint distribution the following rate-region is achievable.

$$\begin{aligned}
R_1 &< I(U;Y_1|AC) + I(\tilde{U}AC;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1) - I(\tilde{V}\tilde{B}\tilde{Y}_2;AC|\tilde{U}\tilde{A}\tilde{C}\tilde{Y}_1)\\
R_2 &< I(V;Y_2|BC) + I(\tilde{V}BC;Y_2|\tilde{B}\tilde{C}\tilde{Y}_2) - I(\tilde{U}\tilde{A}\tilde{Y};BC|\tilde{V}\tilde{B}\tilde{C}\tilde{Y}_2)\\
R_1 + R_2 &< \min\{T_1,T_2,T_3\} + I(U;Y_1|AC) + I(\tilde{U}A;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1C)\\
&\quad - I(\tilde{U}\tilde{A}\tilde{Y}_1;B|\tilde{V}\tilde{B}\tilde{C}\tilde{Y}_2C) + I(V;Y_2|BC) + I(\tilde{V}B;Y_2|\tilde{B}\tilde{C}\tilde{Y}_2C)\\
&\quad - I(\tilde{V}\tilde{B}\tilde{Y}_2;A|\tilde{U}\tilde{A}\tilde{C}\tilde{Y}_1C) - I(A;B|\tilde{U}\tilde{V}\tilde{A}\tilde{B}\tilde{C}\tilde{Y}_1\tilde{Y}_2C) - I(U;V)
\end{aligned}$$

where

$$\begin{aligned}
T_1 &\triangleq I(C;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1) + I(\tilde{V};C|\tilde{B}\tilde{C}\tilde{Y}_2) - I(\tilde{V}\tilde{B}\tilde{Y}_2;C|\tilde{U}\tilde{A}\tilde{C}\tilde{Y}_1)\\
T_2 &\triangleq I(C;Y_2|\tilde{B}\tilde{C}\tilde{Y}_2) + I(\tilde{U};C|\tilde{A}\tilde{C}\tilde{Y}_1) - I(\tilde{U}\tilde{A}\tilde{Y}_1;C|\tilde{V}\tilde{B}\tilde{C}\tilde{Y}_2)\\
T_3 &\triangleq I(C;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1) + I(C;Y_2|\tilde{B}\tilde{C}\tilde{Y}_2) - I(\tilde{V}\tilde{B}\tilde{Y}_2;C|\tilde{U}\tilde{A}\tilde{C}\tilde{Y}_1)\\
&\quad - I(\tilde{U}\tilde{A}\tilde{Y}_1;C|\tilde{V}\tilde{B}\tilde{C}\tilde{Y}_2)
\end{aligned}$$

The rate region is computed with the joint distribution (4).

Dueck's feedback example: We show that Theorem 1 yields the optimal rates for the example described in Section I. Set

$$\begin{gathered}
P_{UV} = P_U P_V, \text{ with } P_U(0) = P_U(1) = P_V(0) = P_V(1) = \frac{1}{2}\\
A = B = \phi, \ P_C(0) = P_C(1) = \frac{1}{2}\\
X : (X_0 = C, X_1 = U, X_2 = V)
\end{gathered} \tag{5}$$

Next, we define the distribution Q that generates C for each block from the variables of the previous block.

$$Q : C = \tilde{Y}_1 \oplus \tilde{U} = \tilde{Y}_2 \oplus \tilde{V} \tag{6}$$

Since $Y_1 \oplus U = Y_2 \oplus V = Z$, the noise variable, the above choice satisfies (3). Finally, substituting (5) in Theorem 1, the rate region is computed as

$$\begin{aligned}
R_1 &< 0 + 1 - 0 = 1\\
R_2 &< 0 + 1 - 0 = 1\\
R_1 + R_2 &< 0 + \min\{T_1, T_2, T_3\} = 2
\end{aligned}$$

IV. Proof Sketch of Theorem 1

Fix a joint distribution as in (2) and pick a conditional distribution $Q_{ABC|\tilde{A}\tilde{B}\tilde{C}\tilde{U}\tilde{V}\tilde{Y}_1\tilde{Y}_2}$ that satisfies (3). Then the joint distribution P over two successive blocks is given by (4).

A. Random code generation

1) *Coding the fresh messages*: Pick $2^{nR_1'}$ $\mathbf{U}$ codewords, each uniformly at random from the typical set $T(P_U)$. Label these codewords $\mathbf{u}(i)$, $i \in \{1, \dots, 2^{nR_1'}\}$. Divide these into 2^{nR_1} equal-sized bins. Similarly, pick a random $\mathbf{V}$ codebook with $2^{nR_2'}$ codewords, each uniformly at random from the set $T(P_V)$. Label these codewords $\mathbf{v}(j)$, $j \in \{1, \dots, 2^{nR_2'}\}$. Divide these codewords into 2^{nR_2} equal-sized bins.
2) *Covering the correlated sources*: In each block, $(\tilde{\mathbf{A}}, \tilde{\mathbf{C}}, \tilde{\mathbf{U}}, \tilde{\mathbf{Y}}_\mathbf{1})$ represents the source sequence to be transmitted to receiver 1, and $(\tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \tilde{\mathbf{V}}, \tilde{\mathbf{Y}}_\mathbf{2})$ the source for receiver 2. The sequences for covering the sources are generated as follows.
   - For each sequence $\tilde{\mathbf{c}} \in \mathcal{C}^n$, pick $2^{n\rho_0}$ $\mathbf{C}$ codewords independently from the conditional typical set $T(P_{C|\tilde{C}})$. Label these codewords $\mathbf{c}(k_0|\tilde{\mathbf{c}})$, $k_0 \in \{1, \dots, 2^{n\rho_0}\}$.
   - For each jointly typical triple $(\tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{u}}, \tilde{\mathbf{y}}_\mathbf{1}, \mathbf{c}(k_0|\tilde{\mathbf{c}}))$, pick $2^{n\rho_1}$ $\mathbf{A}$ sequences, independently from the conditionally typical set $T(P_{A|C\tilde{A}\tilde{C}\tilde{U}\tilde{Y}_1})$. Label these sequences $\mathbf{a}(k_1|k_0, \tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{u}}, \tilde{\mathbf{y}}_\mathbf{1})$, $k_1 \in \{1, \dots, 2^{n\rho_1}\}$.
   - For each jointly typical triple $(\tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{v}}, \tilde{\mathbf{y}}_\mathbf{2}, \mathbf{c}(k_0|\tilde{\mathbf{c}}))$, pick $2^{n\rho_2}$ $\mathbf{B}$ sequences, independently from

the conditionally typical set $T(P_{B|C\tilde{B}\tilde{C}\tilde{V}\tilde{Y}_2})$. Label these sequences $\mathbf{b}(k_2|k_0, \tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{v}}, \tilde{\mathbf{y}}_2)$, $k_2 \in \{1, \ldots, 2^{n\rho_2}\}$.

3) *Channel Input*: For each jointly typical tuple $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}, \mathbf{v})$, pick one input sequence at random $\mathbf{x}$ from the conditionally typical set $T(P_{X|ABCUV})$.

B. Encoding

For block 1, $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{u}}, \tilde{\mathbf{v}}, \tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2)_1$ are chosen a priori at random from $T(P_{ABCUVY_1Y_2})$, and coding starts from step 2 below. For $2 \leq b \leq B$, after receiving $(\mathbf{y_1}, \mathbf{y_2})_{b-1}$ through feedback, the encoder generates the input $\mathbf{x}_b$ for block b as follows.

1) Set $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{u}}, \tilde{\mathbf{v}}, \tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2)_b = (\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}, \mathbf{v}, \mathbf{y_1}, \mathbf{y_2})_{b-1}$.
2) Pick (k_0, k_1, k_2) such that $\mathbf{c}(k_0|\tilde{\mathbf{c}}_b)$, $\mathbf{a}(k_1|k_0, (\tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{u}}, \tilde{\mathbf{y}}_1)_b)$ and $\mathbf{b}(k_2|k_0, (\tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{v}}, \tilde{\mathbf{y}}_2)_b)$ are jointly typical with $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{u}}, \tilde{\mathbf{v}}, \tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2)_b$ according to the joint distribution P in (4). Using techniques similar to [7], [12], one can show that such a triple (k_0, k_1, k_2) can be found if

$$\begin{aligned}
&\rho_0 > I(\tilde{A}\tilde{B}\tilde{U}\tilde{V}\tilde{Y}_1\tilde{Y}_2; C|\tilde{C}) \\
&\rho_0 + \rho_1 > I(\tilde{B}\tilde{V}\tilde{Y}_2; A|\tilde{A}\tilde{C}\tilde{U}\tilde{Y}_1 C) + I(\tilde{A}\tilde{B}\tilde{U}\tilde{V}\tilde{Y}_1\tilde{Y}_2; C|\tilde{C}) \\
&\rho_0 + \rho_2 > I(\tilde{A}\tilde{U}\tilde{Y}_1; B|\tilde{B}\tilde{C}\tilde{V}\tilde{Y}_2 C) + I(\tilde{A}\tilde{B}\tilde{U}\tilde{V}\tilde{Y}_1\tilde{Y}_2; C|\tilde{C}) \\
&\rho_0 + \rho_1 + \rho_2 > I(\tilde{A}\tilde{U}\tilde{Y}_1 A; \tilde{B}\tilde{V}\tilde{Y}_2 B|\tilde{C}C) \\
&\qquad - I(\tilde{A}\tilde{U}\tilde{Y}_1; \tilde{B}\tilde{V}\tilde{Y}_2|\tilde{C}C) + I(\tilde{A}\tilde{B}\tilde{U}\tilde{V}\tilde{Y}_1\tilde{Y}_2; C|\tilde{C}).
\end{aligned} \tag{7}$$

Set $\mathbf{C}_b = \mathbf{c}(k_0|\tilde{\mathbf{c}})$, $\mathbf{A}_b = \mathbf{a}(k_1|k_0, (\tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{u}}, \tilde{\mathbf{y}}_1)_b)$ and $\mathbf{B}_b = \mathbf{b}(k_2|k_0, (\tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{v}}, \tilde{\mathbf{y}}_2)_b)$.

3) If $i \in \{1, \ldots, 2^{nR_1}\}$ and $j \in \{1, \ldots, 2^{nR_2}\}$ are the fresh messages to be transmitted in block b, pick a sequence from bin i of the $\mathbf{U}-$codebook and a sequence from bin j of the $\mathbf{V}-$codebook that are jointly typical according to P_{UV}. Set these as $(\mathbf{u}, \mathbf{v})_b$. As shown in [12], this step is successful if

$$R_1 < R_1', \ R_2 < R_2', \ R_1 + R_2 < R_1' + R_2' - I(U; V). \tag{8}$$

4) As described in Step 3 of Section IV-A, set the channel input $\mathbf{X}_b$ to be the sequence corresponding to the tuple $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}, \mathbf{v})_b$.

No new messages are transmitted in block B; $(\mathbf{u}_B[1], \mathbf{v}_B[1])$ are taken to be the codewords for this block. The reduction in rate due to this is insignificant if B is very large.

C. Decoding

At the end of each block $b\,(2 \leq b \leq B)$, the decoders receive $\mathbf{y}_{1b}$ and $\mathbf{y}_{2b}$, respectively. Each of them then decodes its message corresponding to the *previous* block, as described below.

- *At the end of block* $(b-1)$: Decoder 1 knows the triple $(\mathbf{a}, \mathbf{c}, \mathbf{y_1})_{b-1}$, which is then denoted $(\tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_1)$. Decoder 2 knows $(\mathbf{b}, \mathbf{c}, \mathbf{y_2})_{b-1}$, which is then denoted $(\tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_2)$.
- *Decoder* 1*, at the end of block* b: Upon receiving $\mathbf{y}_{1b}$ at the end of block b, decoder 1 tries to find a triple $(i, k_0, k_1) \in \{1, \ldots, 2^{nR_1'}\} \times \{1, \ldots, 2^{n\rho_0}\} \times \{1, \ldots, 2^{n\rho_1}\}$ such that [2]

$$\begin{aligned}
&(\mathbf{u}(i), \mathbf{a}, \mathbf{c}, \mathbf{y_1})_{b-1} \text{ are jointly typical AND} \\
&\quad \mathbf{c}(k_0|\tilde{\mathbf{c}}), \mathbf{y}_{1b}, \mathbf{a}(k_1|k_0, \mathbf{u}(i), \tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_1) \text{ are jointly typ.}
\end{aligned}$$

If there is a unique such triple (i, k_0, k_1), the bin number of $\mathbf{u}(i)$ is declared to be the message of decoder 1 for block $(b-1)$. Note that decoder 1 now knows $\mathbf{c}_b = \mathbf{c}(k_0|\tilde{\mathbf{c}})$ and $\mathbf{a}_b = \mathbf{a}(k_1|k_0, \mathbf{u}(i), \tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_1)$, which are used for decoding at the end of the next block.

- *Decoder* 2*, at the end of block* b: Upon receiving $\mathbf{y}_{2b}$ at the end of block b, decoder 2 tries to find a triple $(j, k_0, k_1) \in \{1, \ldots, 2^{nR_2'}\} \times \{1, \ldots, 2^{n\rho_0}\} \times \{1, \ldots, 2^{n\rho_2}\}$ such that

$$\begin{aligned}
&(\mathbf{v}(j), \mathbf{b}, \mathbf{c}, \mathbf{y_2})_{b-1} \text{ are jointly typical AND} \\
&\quad \mathbf{c}(k_0|\tilde{\mathbf{c}}), \mathbf{y}_{2b}, \mathbf{b}(k_2|k_0, \mathbf{v}(j), \tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_2) \text{ are jointly typ.}
\end{aligned}$$

If there is a unique such triple (j, k_0, k_1), the bin number of $\mathbf{v}(j)$ is declared to be the message of decoder 2 for block $(b-1)$. Decoder 2 now knows $\mathbf{c}_b = \mathbf{c}(k_0|\tilde{\mathbf{c}})$ and $\mathbf{b}_b = \mathbf{b}(k_2|k_0, \mathbf{u}(i), \tilde{\mathbf{b}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_2)$, which are used for decoding at the end of the next block.

D. Probability of Decoding Error

We can assume that the triple $(i = 1, j = 1, k_0 = 1, k_1 = 1, k_2 = 1)$ was transmitted. The probability of error for decoder 1 for decoding the message of block $(b-1)$ can be expressed as

$$P^1_{(b-1)} = \Pr(E_0) + \Pr(E_1|E_0^c) + \Pr(E_2|E_0^c) \tag{9}$$

where

$$\begin{aligned}
E_0 &\triangleq (\mathbf{u}(1), \mathbf{a}, \mathbf{c}, \mathbf{y_1})_{b-1} \textit{ not} \text{ jointly typical OR} \\
&\quad \mathbf{c}(1|\tilde{\mathbf{c}}), \mathbf{a}(1|1, \mathbf{u}(1), \tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_1), \mathbf{y}_{1b} \textit{ not} \text{ jointly typical} \\
E_1 &\triangleq (\mathbf{u}(i), \mathbf{a}, \mathbf{c}, \mathbf{y_1})_{b-1} \text{ jointly typical AND} \\
&\quad \mathbf{c}(1|\tilde{\mathbf{c}}), \mathbf{a}(k_1|1, \mathbf{u}(i), \tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_1), \mathbf{y}_{1b} \text{ jointly typical for} \\
&\quad \text{some } (i, k_1) \neq (1, 1) \\
E_2 &\triangleq (\mathbf{u}(i), \mathbf{a}, \mathbf{c}, \mathbf{y_1})_{b-1} \text{ jointly typical AND} \\
&\quad \mathbf{c}(k_0|\tilde{\mathbf{c}}), \mathbf{a}(k_1|k_0, \mathbf{u}(i), \tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_1), \mathbf{y}_{1b} \text{ jointly typical for} \\
&\quad \text{some } (i, k_0, k_1) \neq (1, 1, 1)
\end{aligned}$$

Assuming no decoding errors in the previous blocks, $\Pr(E_0) < \epsilon$ due to the random code construction and the property of typical sequences. Averaged over all codebooks, $\Pr(E_1|E_0^c)$ can be upper bounded as follows. Define the list

$$\mathcal{L}_{(b-1)} = \left\{ i : (\mathbf{u}(i), \mathbf{a}, \mathbf{c}, \mathbf{y_1})_{b-1} \text{ jointly typical}, \ 1 \leq i \leq 2^{NR_1'} \right\} \tag{10}$$

This is the list of possible messages at decoder 1 for block $(b-1)$. The list is resolved at the end of block b. In terms of

[2]We emphasize that $(\mathbf{a}, \mathbf{c}, \mathbf{y_1})_{b-1}$ is the same as $(\tilde{\mathbf{a}}, \tilde{\mathbf{c}}, \tilde{\mathbf{y}}_1)_b$. The two notations distinguish the role of the same set of sequences in different blocks.

this list, we can write

$$\Pr(E_1|E_0^c) < E\Big(\sum_{i\in\mathcal{L}_{b-1}}\sum_{k_1=1}^{\rho_1}\Pr\{\mathbf{Y}_{1b},\mathbf{C}(1|\tilde{\mathbf{c}})$$
$$\text{jointly typical with }\tilde{\mathbf{U}}(i),\mathbf{A}(k_1|1,\tilde{\mathbf{U}}(i),\tilde{\mathbf{a}},\tilde{\mathbf{c}},\tilde{\mathbf{y}}_{\mathbf{1}})\}\Big)$$
$$< E\sum_{i\in\mathcal{L}_{b-1}}\sum_{k_1=1}^{\rho_1}2^{-n(I(\tilde{U};C|\tilde{A}\tilde{C}\tilde{Y}_1)-\epsilon)}\cdot 2^{-n(I(\tilde{U}A;Y_1|C\tilde{A}\tilde{C}\tilde{Y}_1)-\epsilon)}$$
$$= E\left(\sum_{i\in\mathcal{L}_{b-1}}1\right)2^{n(\rho_1-I(\tilde{U};C|\tilde{A}\tilde{C}\tilde{Y}_1)-I(\tilde{U}A;Y_1|C\tilde{A}\tilde{C}\tilde{Y}_1)+2\epsilon)}. \tag{11}$$

From the definition (10), we have

$$E\left(\sum_{i\in\mathcal{L}_{b-1}}1\right) < 2^{nR_1'}\cdot 2^{-n(I(U;Y_1|AC)-\epsilon)}. \tag{12}$$

Combining (11) and (12), we see that $\Pr(E_1|E_0^c)$ is upper-bounded by

$$2^{n(R_1'+\rho_1-I(U;Y_1|AC)-I(\tilde{U};C|\tilde{A}\tilde{C}\tilde{Y}_1)-I(\tilde{U}A;Y_1|C\tilde{A}\tilde{C}\tilde{Y}_1)+3\epsilon)}.$$

Hence $\Pr(E_1|E_0^c)$ can be made arbitrarily small if

$$R_1' < I(U;Y_1|AC)+I(\tilde{U};C|\tilde{A}\tilde{C}\tilde{Y}_1)+I(\tilde{U}A;Y_1|C\tilde{A}\tilde{C}\tilde{Y}_1)-\rho_1. \tag{13}$$

Next, we bound $\Pr(E_2|E_0^c)$.

$$\Pr(E_2|E_0^c) < E\Big(\sum_{i\in\mathcal{L}_{b-1}}\sum_{k_0=1}^{\rho_0}\sum_{k_1=1}^{\rho_1}\Pr\{\mathbf{Y}_{1b}\text{ jointly typical}$$
$$\text{with }\mathbf{C}(k_0|\tilde{\mathbf{c}}),\tilde{\mathbf{U}}(i),\mathbf{A}(k_1|k_0,\tilde{\mathbf{U}}(i),\tilde{\mathbf{a}},\tilde{\mathbf{c}},\tilde{\mathbf{y}}_{\mathbf{1}})\}\Big)$$
$$< E\sum_{i\in\mathcal{L}_{b-1}}\sum_{k_0=1}^{\rho_0}2^{-n(I(\tilde{A}\tilde{U}\tilde{Y}_1;C|\tilde{C})-\epsilon)}\sum_{k_1=1}^{\rho_1}2^{-n(I(\tilde{U}AC;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1)-\epsilon)}$$
$$= E\left(\sum_{i\in\mathcal{L}_{b-1}}1\right)2^{n(\rho_0+\rho_1-I(\tilde{A}\tilde{U}\tilde{Y}_1;C|\tilde{C})-I(\tilde{U}AC;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1)+2\epsilon)}. \tag{14}$$

Combining (14) and (12), we can upper-bound $\Pr(E_2|E_0^c)$ as

$$2^{n(R_1'+\rho_0+\rho_1-I(U;Y_1|AC)-I(\tilde{A}\tilde{U}\tilde{Y}_1;C|\tilde{C})-I(\tilde{U}AC;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1)+3\epsilon)}.$$

Hence $\Pr(E_2|E_0^c)$ can be made arbitrarily small if

$$\begin{aligned}R_1' <& I(U;Y_1|AC)+I(\tilde{A}\tilde{U}\tilde{Y}_1;C|\tilde{C})\\ &+I(\tilde{U}AC;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1)-\rho_0-\rho_1.\end{aligned} \tag{15}$$

Thus the probability (9) of decoder 1 incorrectly decoding its message for block $(b-1)$ can be made arbitrarily small if (13) and (15) are satisfied. Similarly, the probability of decoder 2 incorrectly decoding its message can be made small if

$$\begin{aligned}R_2' <& I(V;Y_2|BC)+I(\tilde{V};C|\tilde{B}\tilde{C}\tilde{Y}_2)\\ &+I(\tilde{V}B;Y_2|C\tilde{B}\tilde{C}\tilde{Y}_2)-\rho_2,\end{aligned} \tag{16}$$

$$\begin{aligned}R_2' <& I(V;Y_2|BC)+I(\tilde{B}\tilde{V}\tilde{Y}_2;C|\tilde{C})\\ &+I(\tilde{V}BC;Y_2|\tilde{B}\tilde{C}\tilde{Y}_2)-\rho_0-\rho_2.\end{aligned} \tag{17}$$

Finally, we combine (13),(15),(16),(17) with (7) and use the Fourier-Motzkin technique to eliminate ρ_0,ρ_1,ρ_2. We are left with the following inequalities in terms of R_1' and R_2'.

$$\begin{aligned}&R_1 < I(U;Y_1|AC)+I(\tilde{U}AC;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1)-I(\tilde{V}\tilde{B}\tilde{Y}_2;AC|\tilde{U}\tilde{A}\tilde{C}\tilde{Y}_1)\\ &R_2 < I(V;Y_2|BC)+I(\tilde{V}BC;Y_2|\tilde{B}\tilde{C}\tilde{Y}_2)-I(\tilde{U}\tilde{A}\tilde{Y}_1;BC|\tilde{V}\tilde{B}\tilde{C}\tilde{Y}_2)\\ &R_1+R_2 < \min\{T_1,T_2,T_3\}+I(U;Y_1|AC)+I(\tilde{U}A;Y_1|\tilde{A}\tilde{C}\tilde{Y}_1C)\\ &\quad -I(\tilde{U}\tilde{A}\tilde{Y}_1;B|\tilde{V}\tilde{B}\tilde{C}\tilde{Y}_2C)+I(V;Y_2|BC)+I(\tilde{V}B;Y_2|\tilde{B}\tilde{C}\tilde{Y}_2C)\\ &\quad -I(\tilde{V}\tilde{B}\tilde{Y}_2;A|\tilde{U}\tilde{A}\tilde{C}\tilde{Y}_1C)-I(A;B|\tilde{U}\tilde{V}\tilde{A}\tilde{B}\tilde{C}\tilde{Y}_1\tilde{Y}_2C).\end{aligned} \tag{18}$$

where T_1,T_2 and T_3 are defined in (1). Combining with (8), we obtain the rate region of the theorem.

V. Conclusion

We have derived a single-letter rate region for the two-user broadcast channel with feedback. Using the Marton coding scheme as the starting point, our scheme uses three additional random variables (A,B,C) to cover the correlated information generated at the end of each block. The proposed region can be strictly larger than the no feedback capacity region as shown by Dueck's feedback example. Examples to show how Theorem 1 can be used to compute rates for other broadcast channels (such as the AWGN broadcast channel) will be presented in an extended version of this paper. We can also show that Theorem 1 reduces to the no-feedback capacity region for a physically degraded broadcast channel, consistent with the result of [3]. Investigating if one can similarly exploit correlated information in broadcast channels with partial or noisy feedback is part of future work.

References

[1] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform Theory*, vol. 25, no. 3, pp. 306–311, 1979.

[2] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, 1998.

[3] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. 24, May 1980.

[4] G. Dueck, "Partial feedback for two-way and broadcast channels," *Inform. and Control*, vol. 46, pp. 1–15, July 1980.

[5] L. H. Ozarow and S. K. Leung-Yan-Cheong, "An achievable region and outer bound for the gaussian broadcast channel with feedback," *IEEE Trans. on Inform. Theory*, vol. 30, no. 4, pp. 667–671, 1984.

[6] S. R. Bhaskaran, "Gaussian broadcast channel with feedback," *IEEE Trans. on Inform. Theory*, vol. 54, no. 11, pp. 5252–5257, 2008.

[7] T. S. Han and M. H. M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inform, Theory*, vol. IT-33, no. 5, pp. 641–650, 1987.

[8] G. Kramer and C. Nair, "Comments on 'Broadcast channels with arbitrarily correlated sources'," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2009.

[9] P. Minero and Y.-H. Kim, "Correlated sources over broadcast channels," *CoRR:http://arxiv.org/abs/0902.1591*, vol. abs/0902.1591, 2009.

[10] S. Choi and S. S. Pradhan, "A graph-based framework for transmission of correlated sources over broadcast channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2841–2856, 2008.

[11] W. Kang and G. Kramer, "Broadcast channel with degraded source random variables and receiver side information," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2008.

[12] A. El Gamal and E. C. van der Muelen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, 1981.

[13] T. S. Han, "A general coding scheme for the two-way channel," *IEEE Trans. Inform. Theory*, vol. 30, no. 1, pp. 35–43, 1984.